TO: Stormwater Tech Team

FROM: Karen Tarnow and Mike Poulsen, DEQ

DATE: 9/18/08

RE: Proposed method for calculating representative LU concentrations

We have reviewed Carl's memo and Kristine's comments and pondered this long and hard. We also had really good one-on-one conversations with Kristine and with Dawn/Jason that did a LOT to enlighten us - it's easy to see how our conference calls were unfortunately not the most effective vehicle for creating mutual understanding in this instance.

So, we're putting this out there in hopes that it may help us reach consensus on how to proceed.

Here's how we see it. The difference in the approaches we've been debating comes down to a difference in conceptual models. Below we've described our "take" on each model and some additional thoughts on each. We also offer a third approach which is a modified version of the second model that we think will better represent variability.

Option 1
One conceptual model assumes that the sampled locations are representative of the types of industries within this LU category, and thus the data we collected are representative of the types/range of concentrations. Therefore, the values can be averaged to develop a representative concentration for this LU.

o With this approach, you need to average the values from each location to calculate the representative concentration, because if you used all 20 data points you have the problem of having different numbers of samples from different basins which skews the results.

Option 2
The other conceptual model assumes that larger basins will encompass a greater diversity of activities for this LU as well as a greater mix of pervious/impervious surfaces. Since the runoff from the site integrates the volume and concentration of runoff from each unique activity area, the resulting concentration will be more representative of the central tendency of concentrations from all sites within this LU. Therefore, the data from each sampled location should be weighted proportional to the flow from that basin before averaging the results when calculating the representative concentration for this LU.

o Kristine's concern is that this approach is not valid unless we can demonstrate a correlation between flow and concentration. This concern may be based upon a misunderstanding about the premise of the approach. The premise behind "weighting" the data is based upon the assumption that the concentration of larger basins will be more *representative* of the central tendency of concentrations from <u>all</u> sites of this LU because the runoff from these sites already "averages" the runoff from a mix of activities and pervious/impervious surfaces. Thus, the relationship is not between flow and concentration, but between flow and representativeness of concentration. We use flow in the equation rather than area because flow captures the variable of perviousness in addition to area.

o One negative of this approach is that we greatly reduce the sample size because it is necessary to average the data points from each site. This may be unavoidable.

o Another concern we have with the approach presented in the 9/2/08 memorandum is that the flow-weighted concentration for each basin (Column G) does not have a meaningful relationship to the actual range of concentrations measured. For this reason we are proposing a modified approach – Option 3.

Option 3
For a modified approach, we propose to use the flow-weighted mean concentration as in Option 2, although calculated as demonstrated in Column H. For the range of concentrations, we propose to use the range of average concentrations for each basin (Column E). This is not a rigorous statistical measure of variability, but it can be used to give us some sense of the range of possible concentrations. This method has the same limitation that there will be a small number of data points. However, this appears to be an unavoidable consequence of our sampling program.

The following are a number of "truisms" (not criticisms) to keep in mind as we strive to reach resolution on this issue.
1. Stormwater data are inherently variable
2. Our data set is small, especially when you consider the variability
3. Sample locations were not selected based upon a probabilistic approach
4. The SAP was not based on a solid statistical design
5. We don't have any way of determining the "correct" answer
6. The LU concentrations are intended to be used in a modeling exercise that will produce very broad, general insights into the effects of stormwater on the harbor

We have attached a spreadsheet that shows how we envision the calculations involved in each approach.

If we go forward with more than one approach, we need to agree upfront about how we will use the various numbers once we have them lest we create an opportunity for introducing bias into the decision-making after the results of the calculations become available.

The simplest approach would be to agree to use one method. **In our opinion, the flow-weighted approach would better characterize the appropriate LU concentration.** We suspect that the results will be significantly influenced by a lot of randomness in the data set and therefore we wouldn't expect to see huge differences in mean values across the board by using one approach or the other.

To evaluate variability**, we propose using the approach in Option 3** and looking at ranges of average concentrations in test basins. We'll need to decide HOW we use these values – so far starters we'll propose that we use percentiles (e.g., 25th and 75th) to quantify the uncertainty in concentrations.